A Heuristic Intermolecular Potential Function for Formaldehyde–Water Based on ab Initio Molecular Orbital Calculations

S. Swaminathan, R. J. Whitehead, Elias Guth, and David L. Beveridge*

Contribution from the Chemistry Department, Hunter College of the City University of New York, New York, New York 10021. *Received January 10, 1977*

Abstract: A systematic procedure for the determination of intermolecular potential functions from quantum mechanical calculations of pairwise interaction energies is described and applied to the formaldehyde–water system. The data base and certain other variables evolve in the course of the calculation in order to produce an acceptable function with a minimum of calculated grid points. Standard deviations representative of the tolerance and predictive value of the function are generated. Contour maps of energy vs. orientationally optimized geometries are reported and discussed in terms of the formaldehyde–water interaction and the hydration of a carbonyl group.

I. Introduction

We describe herein the development and application of methodology for determining potential functions representative of ab initio quantum mechanical calculations of intermolecular interaction energies. Our goal is to establish a systematic and well-characterized general approach to obtaining the best quality function in the most economical way possible. To accomplish this, certain variables in the procedure will be permitted to evolve in the course of determining a particular function, leading to a heuristic method. The functions obtained are intended for use in computer simulation studies of molecular assemblies. The initial study involves the determination of a pairwise potential function for the formaldehyde–water interaction.

The following section of this paper provides background material for the project. The method of procedure is given in detail in section III, including considerations on specification of variables and computational technology. The calculations are described in section IV, followed in section V by a presentation of results and discussion thereof. Summary and conclusions are collected in section VI.

II. Background

Theoretical studies of chemical systems using statistical mechanics and molecular dynamics require rapid, accurate evaluation of the configurational energy of the system or the force at various positions within the system. In numerical calculations at present this is accomplished using a potential function, a simple analytical expression for the energy of the system as a function of configurational coordinates and disposable parameters. Forces can be determined from the gradient of the potential function. The quality of the numerical results obtained in computer simulation studies on chemical systems depends directly on the quality of the potential functions used.

The simplest way in which potential functions are used in numerical calculations is with the assumption of pairwise additivity, wherein the total configurational energy E of an assembly is written as a sum of pairwise interaction energies between the individual particles of the assembly.

$$E(X^N) = \sum_{A<B} V_{AB} \tag{1}$$

where X^N represents the configurational coordinates of N particles and V_{AB} is a pairwise potential function for molecules A and B. The most widely known pairwise potential function is the Lennard-Jones 6–12 potential function,[1]

$$V_{AB}(r) = 4\epsilon_{AB}[(\sigma_{AB}/r)^{12} - (\sigma_{AB}/r)^6] \tag{2}$$

where r is the interparticle separation and ϵ_{AB} and σ_{AB} are disposable parameters characteristic of the interaction of molecules A and B. Numerous other functional forms for pairwise potential functions have been used in various applications.[2] Generally the true intermolecular energy is unknown, so the functional form for the pairwise interaction is chosen with a physical model in mind, and the disposable parameters are determined empirically or semiempirically from independent considerations or experimental data. This type of potential function will be henceforth referred to as an empirical potential function (EPF).

Currently, good estimates of the true intermolecular pairwise interaction energy can be obtained using molecular quantum mechanics. Recent advances in computer technology and the development of efficient computational algorithms have made ab initio quantum mechanical calculations tractable for large molecules and molecular complexes. However, at the present moment the computation time for individual pairwise interaction energies still precludes their direct use in statistical thermodynamic or molecular dynamics calculations.

One currently viable approach to the problem is to use quantum mechanics to generate a data base of interaction energies at various points in configuration space and choose the disposable parameters in an analytical function against the data base in a least-squares sense. Thus one obtains an "analytical potential function" (APF) representative of ab initio quantum mechanical pairwise interaction energies which can be used in computer simulations in the usual manner. Clementi and co-workers have recently pioneered this approach, and their potential functions representative of large basis set ab initio molecular orbital calculations for ion–water[3] and water–water interactions[4,5] have been used in several recent statistical thermodynamic[6] and molecular dynamic[7] studies of water and solution structure, including studies of the thermodynamic indices of ion hydration from this laboratory.[8,9]

With the feasibility of generating APF's representative of quantum mechanical calculations thus established, there remains to consider the most economical way to proceed. A number of details must be specified in establishing a general procedure for determining APF's including (a) quality of points in the data base, (b) functional form for the APF, (c) size of the data base, (d) strategy for choosing configurational coordinates of points in the date base, (e) weighting of points in the curve fitting procedure, and (f) the predictive value of the function. Considering these factors, we have recently developed methodology whereby decisions on variables are evolved in the course of the determination of the potential function for a given intermolecular interaction. Thus intrinsic in the procedure is a discovery of the optimum specification of the variables, thus involving a heuristic method leading to a "heuristic potential function" (HPF).

Before passing to details of our procedure it is necessary to comment further on pairwise additivity. While this approximation holds fairly rigorously for simple nonpolar particles, for a preponderant number of systems of chemical interest at least three-body terms will be significant, if not those of even higher order. The HPF procedure can in principle accommodate higher order effects. Initial studies, however, will be carried out for pairwise potential functions with close attention to the approximations involved in their application to chemical problems. In spite of the problems therein, computer simulation research on molecular systems is currently proceeding on the basis of pairwise or effective pairwise[10] additivity. Evidence of the viability of this approach comes from the excellent agreement between observed oxygen-oxygen pair correlation function for liquid water and that computed by Clementi et al.[5] using a pairwise additive APF in a Monte Carlo procedure.

III. Methodology

This section includes a detailed description of the HPF methodology followed by general considerations on specification of variables and computational details.

The HPF Method. Consider the problem as the determination of a particular intermolecular potential function representative of a given level of quantum mechanical interaction energies as economically as possible. The other variables in the procedure are related to the data base and the curve fitting procedure. The data base $\boldsymbol{B}$ consists of a set of energies calculated at various points in the configuration space of the problem,

$$\boldsymbol{B} = \{\boldsymbol{R},\boldsymbol{E}\} \tag{3}$$

where $\boldsymbol{R}$ is the set of entire geometrical parameters (coordinates or interatomic distances) corresponding to the energies $\boldsymbol{E}$. The specification of data base variables involves determining the number and position of points to be included. The number and disposition of points should be large and diverse enough to be representative of $E(\boldsymbol{R})$ but small enough to be economical. The curve fitting variables include the tolerance or standard deviation to be required of the potential function and any strategy involving preferential weighting of points in the data base. The tolerance should be low enough to enable, for example, calculation of internal energy at the order of experimental error of thermodynamic measurements on chemical systems.

The HPF method involves a dynamic decision making process on specification of certain data base and curve fitting variables. The procedure can be best illustrated with data base variables. We begin with an initial data base $\boldsymbol{B}$ consisting of N_1 points, N_1 depending of course on the size of the molecules and the extent of configuration space to be spanned. These N_1 points in configuration space are chosen on the basis of some strategy; various alternatives will be discussed below. The energies $\boldsymbol{E}$ are calculated quantum mechanically for these positions, giving $\boldsymbol{B}_1$. An analytical functional form V is adopted (see also below) and fit to the initial data base by minimizing the sum of squares of the deviations Δ_1, where

$$\Delta_1 = \frac{1}{N_1}\sum_{i}^{N_1} [E_i(\boldsymbol{R}_i) - V_i(\boldsymbol{R}_i, \boldsymbol{a}, \boldsymbol{b}, \ldots)]^2 \tag{4}$$

with $\boldsymbol{a}$, $\boldsymbol{b}$, ..., being symbols for sets of disposable parameters. The result is the analytical potential function

$$V_1(\boldsymbol{R},\boldsymbol{a},\boldsymbol{b}, \ldots) \tag{5}$$

where $\boldsymbol{a}$, $\boldsymbol{b}$, ..., give the best fit of the function to the data base in a least-squares sense. The statistics on the curve fitting procedure are fully generated, including standard deviation σ_1, residuals, etc., and tell us how well the APF fits the data base. At this point we are also interested in a more difficult question: how well the data base itself represents true energy in the configuration space. To estimate this we generate a "test" data base $\boldsymbol{B}_1' = \{\boldsymbol{R}_1',\boldsymbol{E}_1'\}$ consisting of N_1 points outside the initial data base. The predictive value of the function is tested by using the function to calculate the points in $\boldsymbol{B}_1'$ and comparing them with their quantum mechanically calculated counterparts, $\boldsymbol{E}_1'$. The resulting standard deviation σ_1' is a test of the *predictive* value of the V_1.

The statistics σ_1 and σ_1' together indicate the quality of V_1. These can be compared against a prescribed desired tolerance for the function. Initially the tolerance is unlikely to be satisfied and steps must be taken to improve the quality of the function. The most direct way is to increase the size of the data base. Since we have N_1' additional points already calculated for test purposes, the easiest way to do this is to concatenate the initial data base and the test data base. This gives a data base N_2 of a size $N_1 + N_1'$, for which we can redetermine the APF. The result of this step is V_2 ($\boldsymbol{R}$, $\boldsymbol{a}$, $\boldsymbol{b}$, ...) with a standard deviation σ_2 and predictive value on a new test data base $\boldsymbol{B}_2$ of σ_2'. This procedure is repeated until an APF within tolerance is obtained, and for which the statistics are stable on successive iterations. Thus in the course of determining the potential function we have evolved a data base representative enough of energy in configuration space to produce a potential function within prescribed tolerances. A straightforward extension of this procedure allows curve fitting variables as well as data base variables to evolve in the procedure.

Specification of Variables. The totality of variables involved in the procedure can be conveniently grouped into "static" and "dynamic" variables, depending upon whether a variable is fixed or evolves in the course of an HPF determination. The static variables are the analytical form adopted for the potential function and the quality of quantum mechnically calculated points in the data base. The dynamic variables are certain data base variables and curve fitting variables.

The data base variables in the HPF procedure are (a) the size of the initial data base N_1, (b) the size of the test sets N_i, and (c) the position of the points in configuration space. Note that N_1 and N_i are static data base variables whereas the total number of points in the data base, N, is a dynamic variable. N is given by the expression

$$N = N_1 + \sum_{i=1}^{m-1} N_i' \tag{6}$$

where m is the number of cycles to obtain stability and statistics within tolerance in the procedure. The size of the initial data base will depend on the volume of configuration space under consideration. In considering alternative strategies for positioning initial data base points in configuration space, we can identify two extremes: random sampling and grid search. In random sampling the points are chosen by a Monte Carlo technique using a random number generator. Grid search involves laying a multidimensional grid on configuration space and taking data base points at each intersection. Either of these two procedures could be modified by importance sampling, concentrating data base points in regions of special interest. Our initial efforts as described herein are based on a random sampling strategy.

The curve fitting procedure also involves several aspects, some static and some dynamic with respect to the determination of the HPF. The curve fitting method itself is a multidimensional nonlinear least-squares procedure, and currently we are using an algorithm based on Powell's method of isoconjugate directions.[11] The dynamical aspects of the curve fitting procedure are involved with preferential weighting of certain points. This can be keyed on the requirements of the potential function in the particular type of calculation under consideration. For example, in the numerical calculation of a configurational partition function Z for statistical thermodynamics, where

$$Z = \frac{1}{N}\sum_i e^{-E_i/kT} \qquad (7)$$

we can rearrange this expression to

$$Z = \frac{1}{N} e^{-E_0/kT} \sum_i e^{-\Delta E_i/kT} \qquad (8)$$

where $\Delta E = E_i - E_0$ and E_0 is the minimum energy included in the summation. Because of the behavior of a decaying exponential, only values of ΔE below a certain threshold will make a significant contribution to Z. This suggests tailoring a potential function to describe low energy regions of the system especially well. This can be accomplished by concentrating data base points in low energy regions (importance sampling) or by weighting the low energy points in the data base heavier in the curve fitting procedure. Analogous considerations would apply in the Monte Carlo calculation of average quantities using the Metropolis method.[12]

We currently favor preferential weighting of points for statistical thermodynamic potential functions, because while higher energy points need not be described accurately by the potential function they must be kept above threshold. Thus a refined weighting procedure can be based on a minimum standard deviation for points below threshold while describing high energy points anywhere above threshold. The calculations described herein involve a weight function of the form $1 + \alpha e^{-\Delta E_i/kT}$ in the curve fitting procedure minimizing the mean square deviation

$$\Delta = \frac{1}{N}\sum_i [E_i(\boldsymbol{R}_i) - V_i(\boldsymbol{R}_i, \boldsymbol{a}, \boldsymbol{b}, \ldots)]^2(1 + \alpha e^{-\Delta E_i/kT}) \qquad (9)$$

Here the weighting parameter α may also evolve dynamically in the course of determining the HPF or be preset on the basis of exploratory calculations. The standard deviation of points below threshold $\sigma_<$ and the frequency of errors wherein a point above threshold is predicted to be below threshold is computed for each cycle. The parameter α can be adjusted in small increments to minimize $\sigma_<$ and ν simultaneously in the HPF procedure. Thus we obtain an HPF precisely tailored for the calculation of statistical thermodynamic partition functions.

The remaining variables in the procedure to discuss are the functional form of the potential functions generated in the HPF procedure and the quality of points in the data base. The functional form considered in this study should have three attributes: (a) it must be straightforwardly generalizable to handle any intermolecular interaction, (b) it must be easily generalizable to include additional parameters, and (c) the parameters must be identified with clearly defined features of the molecular structure (atoms, pseudoatoms, functional groups, etc.) so as to provide for the possibility of the transfer of disposable parameters from one potential function to another. A simple form of this type for the pairwise case is

$$V_{AB} = \sum_{i<j} [a_i a_j/r_{ij}^{12} + \sum_n (b_{in} b_{jn}/r_{ij}^{n})] \; i \in A, j \in B \qquad (10)$$

where the i's and j's are atomic or pseudoatomic centers on molecules A and B, respectively, and r_{ij} is the intercenter distance. The parameters a_i in the essentially "hard" part and b_i in in the essentially "soft" part of the function are determined in a least-squares sense. The number of parameters is controlled by the summation over n. While our procedure does not require any physical significance for the parameters, the powers of r are reminiscent of a multipole expansion of the energy between atomic centers. Variations on this function are easily formulated, such as using an exponential decay in the repulsive part.

The quality of points in the data base is determined by level of quantum mechanical calculation effected for the calculation of energies. The HPF procedure is of course independent of the method used for energy evaluation and thus the entire range of empirical, semiempirical, and nonempirical methods of molecular quantum mechanics are possible alternatives. Our principal efforts will be at the nonempirical level.

Computational Details. The HPF method as described above can be automated to a high degree using the sophisticated operating systems of 3d-generation digital computer equipment. The principal digital computer programs involved are the quantum mechanical programs for energy calculations and a set of programs collectively referred to as "HPF" which manages data base generation, curve fitting, and the overall flow of computational events. Present versions of programs are prepared for the CUNY IBM 370-168 system. The HPF program uses the IBM system utility ISDRVR together with "dependent job processing" to submit ordered jobs to the input stream. By linking the jobs submitted, HPF can transfer control to other programs for energy evaluation and then be automatically restarted after the energies are available. Intermediate results are frequently inspected to assure that unanticipated computational problems have not occurred.

IV. Calculations

As an initial application we have chosen the formaldehyde-water system. The formaldehyde-water system is a prototype for the hydration of a carbonyl group and is thus of both chemical and biochemical interest. Previous theoretical studies of this system are due to Iwata and Morokuma,[13] Johansson and Kollman,[14] and Del Bene.[15]

The data base for the HPF determination is most conveniently specified in terms of the geometry of water relative to formaldehyde, with the internal coordinates of each moiety held fixed. All geometries were generated by random deployment of positional and orientational coordinates of water within 5.5 Å of the center of mass of formaldehyde molecule. This distance represents twice the average diameter of a water molecule and is intended to produce a function capable of describing the formaldehyde-water interaction in the region of the first and second hydration shells. The choice of a functional form (see below) with proper long range limiting behavior effectively extends the range of the potential function to all configuration space within good approximation. The initial data base for the calculation consisted of 100 points and all test sets involved 25 points each.

The formaldehyde-water (F-W) intermolecular interaction energies were calculated using matrix Hartree-Fock methods incorporating an STO-6-31G basis set on each atom. The 6-31G basis was developed by Pople et al.[16] and is an extended set with split valence shell functions. It was selected for calculations herein as the largest basis feasible within local economic constraints. The effects of basis set distortion were neglected. The intramolecular geometries for both moieties were maintained at their available experimental values in all calculations. The intramolecular geometrical parameters for formaldehyde were taken as $r_{CO} = 1.203$ Å, $r_{CH} = 1.101$ Å, ∠HCO = 121.75°, and ∠HCH = 116.5°. For water we use $r_{OH} = 0.958$ Å and ∠HOH = 105°. The computer time required for obtaining each individual formaldehyde-water energy was approximately 4 min using the program Gaussian-70 on the IBM 370-168.

The HPF procedure was carried out on this data base using two different functional forms, both adaptations of the general form given in eq 10. In both cases parameters were identified with each of the atoms of the molecules. In addition, two

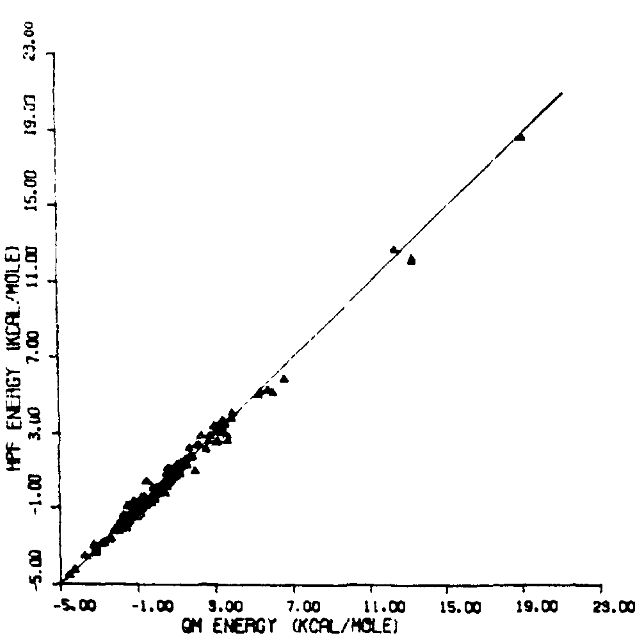

Figure 1. Formaldehyde-water interaction energies calculated from the (12-1-3) HPF vs. corresponding quantum mechanically calculated values.

pseudoatomic centers were introduced for each molecule, coordinated to the oxygen atoms. These type of parameters have proved advantageous in previous empirical and analytical potential functions for water-water interactions. In each case these pseudocenters were tetrahedrally coordinated to the oxygen atom in water and trigonally coordinated to the oxygen in formaldehyde at a distance of 0.1852 Å. The explicit forms of the two functional forms under consideration are

$$V_{12\text{-}1} = \sum_{i<j} \{a_i a_j / r_{ij}{}^{12} + b_i b_j / r_{ij}\} \qquad i \in F, j \in W \quad (11)$$

henceforth referred to as the (12-1) potential, and

$$V_{12\text{-}1\text{-}3} = \sum_{i<j} \{a_i a_j / r_{ij}{}^{12} + b_i b_j / r_{ij} + c_i c_j / r_{ij}{}^{3}\} \qquad i \in F, j \in W \quad (12)$$

hereafter referred to as the (12-1-3) potential. The summations extend over all atoms and pseudoatoms of each molecule. As we ultimately plan to use these functions in Monte Carlo simulation studies of aqueous solutions of formaldehyde-water, the weighting procedure described in eq 9 was employed. Exploratory calculation showed that best results were obtained with $\alpha = 100$ and this value of α was used for all calculations described herein.

V. Results and Discussion

The results of HPF determinations of (12-1-3) and 12-1) potentials are discussed in turn below. The data base $\boldsymbol{B}$ was the same for both determinations. The statistics $\sigma_<$, $\sigma_<'$, and ν are reported for each HPF cycle and a plot of calculated vs. observed energies is included to ascertain the quality of the potential function obtained. The final values of parameters are listed for the (12-1-3) function. As an additional check on the calculation and preliminary application of the results, a sampling of calculated interaction energies as a function of geometry is presented for two two-dimensional slices of the formaldehyde-water interaction in Cartesian space.

(12-1-3) Potential. The statistics of the HPF determination of a potential function for formaldehyde-water based on the (12-1-3) form are given in Table I. For the initial data base of 100 points, the standard deviation of the function was 0.22 kcal/mol. The standard deviation of the values predicted by the function against the corresponding quantum mechanically calculated values for 25 randomly chosen test points was 0.68 kcal/mol. The calculation was continued for four more cycles. The standard deviation of the successive total samples remained well within acceptable limits. The standard deviations for test sets were always larger than the initial sets as expected. On the third cycle the description of the test set degenerated slightly, indicating the sampling of a region of configuration space not well described by the function. On the following cycles the statistics for both the initial set and the test set are back within acceptable limits, confirming the self-correcting nature of the HPF procedure.

Table I. Statistics σ and ν in the Determination of a (12-1-3) HPF Representative of SCF-MO 6-31G Calculations for Formaldehyde-Water

HPF cycle i	Points in data base N	Initial sample $\sigma_<$	Initial sample ν	Test sample $\sigma_<$	Test sample ν
1	100	0.22	0.0	0.68	0.0
2	125	0.23	0.0	0.28	0.0
3	150	0.20	0.0	1.15	0.04
4	175	0.23	0.006	0.58	0.0
5	200	0.23	0.005	0.36	0.0

The final curve fitting of the 200 data points resulted in a standard deviation of 0.23 kcal/mol. The final values for the disposable parameters in the potential function are, in atomic units,

H_2CO $\quad a_C = -5.40 \times 10^{6} \qquad a_O = -1.05 \times 10^{6}$
$b_C = -2.01 \times 10^{-1} \qquad b_O = 3.75$
$c_C = -1.35 \times 10^{1} \qquad c_O = 1.93 \times 10^{1}$
$a_H = -3.09 \times 10^{5} \qquad a_L = -4.06 \times 10^{5}$
$b_H = 2.54 \times 10^{-1} \qquad b_L = -2.08$
$c_H = 1.84 \qquad c_L = -9.41$
H_2O $\quad a_O = -6.33 \times 10^{-1} \qquad a_H = -9.59 \times 10^{-4}$
$b_O = -2.06 \qquad b_H = 3.93 \times 10^{-1}$
$c_O = 9.71 \times 10^{-1} \qquad c_H = -1.15 \times 10^{-1}$
$a_L = 1.20 \times 10^{-1}$
$b_L = 6.37 \times 10^{-1}$
$c_L = 3.50 \times 10^{-1}$

where C, O, H, and L refer to the carbon, oxygen, hydrogen, and lone pair center, respectively. The nature of the procedure does not permit discussing the physical significance of any of the individual parameters; there is even no claim for uniqueness. The function is simply analytical and describes the intraction according to the statistics as reported.

A further indication of the quality of the function is found from a plot of the values predicted each of the 225 energies in the data base vs. their quantum mechanically calculated value. This plot for the (12-1-3) function is shown in Figure 1. Perfect agreement would be a straight line of unit slope; the scatter about this line is a measure of the quality of the function. The standard deviation is of course a composite measure of scatter, but the plot carries additional information on the function. The effect of preferential weighting of low energy points is clearly evident in that the scatter about low energy values is smaller than for high energy values. The geometries poorly described by the function are all in high energy regions and would not contribute significantly to a statistical thermodynamic partition function.

As a final test as well as an initial application of the function, the nature of the formaldehyde-water potential was investigated by generating contour diagrams of energy as a function of intermolecular geometry. The search was limited to two regions which were felt to be of structural chemical interest:

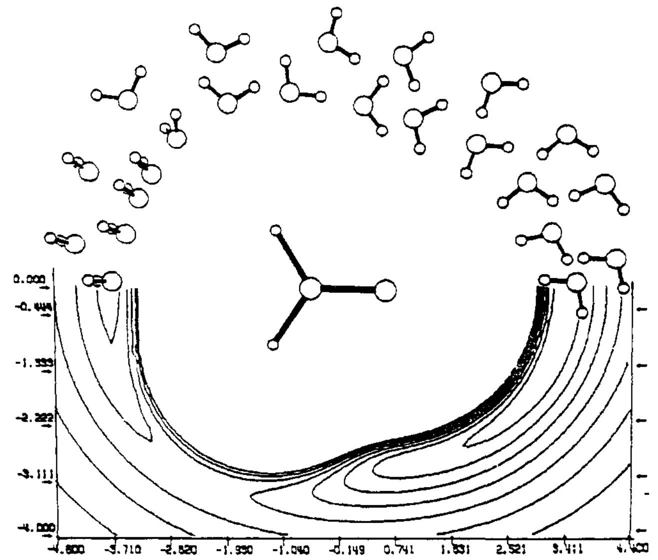

Figure 2. Isoenergy contour map of orientationally optimized formaldehyde-water pairwise interaction energies calculated from the (12-1-3) HPF for the formaldehyde molecular plane. The minimum energy contour line is at −5.4269 kcal/mol with successive lines spaced at an interval of 0.6 kcal/mol. The distance coordinates refer to the separation between the centers of mass of formaldehyde and water in Å. The molecular geometry corresponding to a given energy is depicted in a mirror image position in the top half of the figure. The relative size of the formaldehyde and water molecule structures are scaled down to make the figure more legible.

Table II. Statistics σ and ν in the Determination of a (12-1) HPF Representative of SCF-MO 6-31G Calculations for Formaldehyde-Water

HPF cycle i	Points in data base N	Initial sample $\sigma_<$	Initial sample ν	Test sample $\sigma_<$	Test sample ν
1	100	0.47	0.0	0.51	0.0
2	125	0.47	0.0	0.43	0.0
3	150	0.37	0.0	0.41	0.04
4	175	0.37	0.006	1.34	0.0
5	200	0.44	0.005	0.45	0.0

one in which the center of mass of the water was constrained to be in the plane of the formaldehyde and the other in which the center of mass of the water was perpendicular to the molecular plane and coincident with the plane of the carbonyl group. The energy of each particular formaldehyde-water positional configuration was minimized with respect to orientation of the water molecule. The resulting isoenergetic maps are shown in the lower halves of Figures 2 and 3 and the corresponding orientations of the water molecule are shown in the upper portion of these figures. There are two minima evident in Figure 2, one stabilized by a C=O···H hydrogen bond and the consistent with dipole-dipole stabilization. The plot demonstrates the function to be accommodating the limiting behavior of the interactions as well as vicinal geometries smoothly.

(12-1) Potential. Our initial HPF determination was carried out on the (12-1) form of the potential function. This function ultimately proved unsatisfactory, and the details are instructive with regard to understanding the limitations inherent in HPF and related methodology. The statistics on the HPF cycles in the (12-1) determination and the comparison of quantum mechanically calculated and HPF energies are given in Table II and Figure 4, respectively. The results show the same general characteristics as described above for the (12-1-3) determination, with slightly higher standard deviations in Table II and scatter in Figure 4. The statistics do not indicate any serious deficiencies in the function.

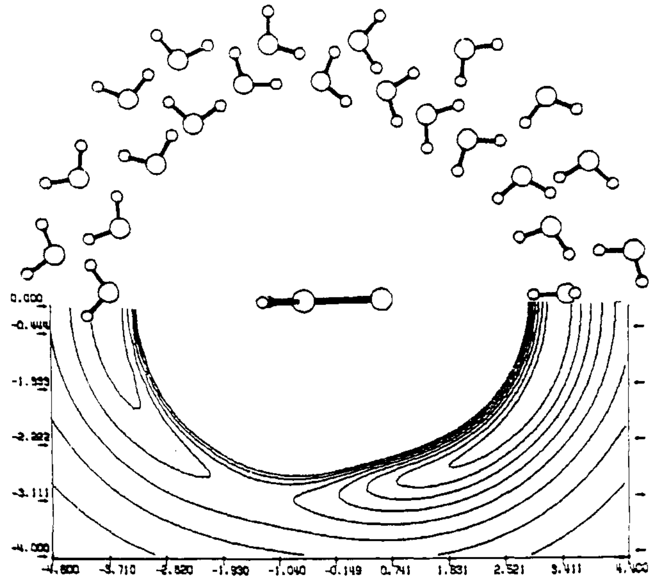

Figure 3. Isoenergy contour map of orientationally optimized formaldehyde-water pairwise interaction energies for the plane coincident with the C-O bond and perpendicular to the molecular plane. See caption to Figure 2 for other details.

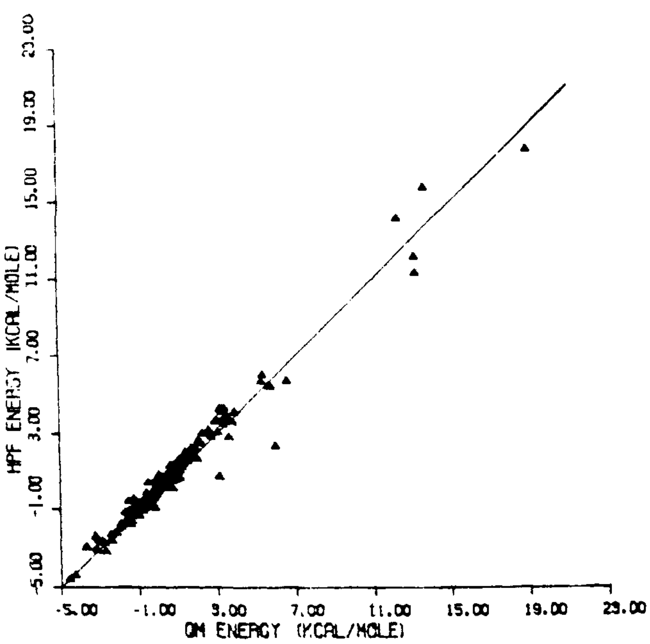

Figure 4. Formaldehyde-water interaction energies calculated from the (12-1) HPF vs. corresponding quantum mechanically calculated values.

Examination of the function on the basis of the basis of the orientationally optimized isoenergetic contour maps revealed a problem. The in plane map, for the (12-1) function, analogous to Figure 2 for the (12-1-3) function, is shown in Figure 5. Encircled on the plot is a distinctly unphysical region, indicating an inadequacy of the function in this region. This problem, while not significant enough to influence the results of a simulation, was clearly rectified in the improved fit gained in the (12-1-3) function. Thus it is essential in assessing the quality of an HPF to consider information from all possible

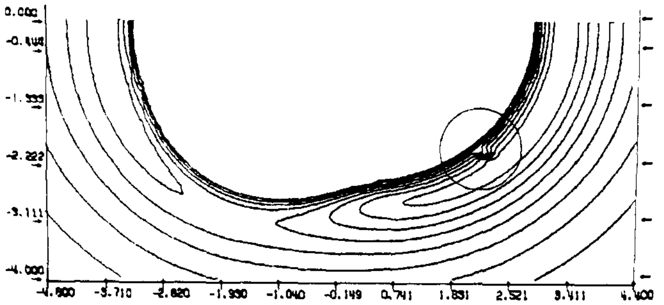

Figure 5. Isoenergy contour map of orientationally optimized formaldehyde water pairwise interactions analogous to Figure 2 but based on the (12-1) HPF. The minimum energy contour is −5.7079 kcal/mol with successive lines spaced at an interval of 0.6 kcal/mol.

sources: magnitudes of the statistics, stability of the statistics, and direct examination of results produced by the function.

VI. Summary and Conclusions

A heuristic method has been described for the determination of intermolecular potential functions from quantum mechanical calculations and applied to the formaldehyde-water system. The stability of statistical indices of the function and the reasonable computed energy grids indicate that the procedure works satisfactorily and within acceptable tolerance. The fact that satisfactory numerical results were obtained for a relatively small data base with respect to the positional and orientational degrees of freedom in the problem indicates that the heuristic potential function approach can be a viable means of generating intermolecular potential functions for larger polyatomic moieties and thus provide a basis for computer simulation studies of molecular assemblies of chemical and biochemical interest.

Acknowledgment. This research was supported in part by NIH Grant 1-R01-NS12149-01 from the National Institute of Neurological and Communicative Diseases and Stroke and a CUNY Faculty Research Award. D.L.B. acknowledges a U.S. Public Health Service Research Career Development Award 6TK04-GM21281 from the National Institute of General Medical Studies. This project developed from discussions with Professors J. W. McIver and G. W. Schnuelle and their helpful contributions are gratefully acknowledged.

References and Notes

(1) J. O. Hirshfelder, C. F. Curtiss, and R. B. Bird, "Molecular Theory of Gases and Liquids", Wiley, New York, N.Y., 1954.
(2) F. Kohler, "The Liquid State", Verlag Chemie, Weinheim/Bergstr., Germany, 1972, Chapter 5.
(3) H. Kistenmacher, H. Popkie, and E. Clementi, *J. Chem. Phys.*, **59**, 5842 (1973).
(4) H. Popkie, H. Kistenmacher, and E. Clementi, *J. Chem. Phys.*, **59**, 1325 (1973).
(5) G. C. Lie, E. Clementi, and M. Yoshimine, *J. Chem. Phys.*, **64**, 2314 (1976).
(6) R. O. Watts, *Mol. Phys.*, **28**, 1069 (1974).
(7) R. O. Watts, *Mol. Phys.*, **32**, 659 (1976).
(8) G. W. Schnuelle and D. L. Beveridge, *J. Phys. Chem.*, **79**, 2566 (1975).
(9) G. W. Schnuelle, S. Swaminathan, and D. L. Beveridge, *Theor. Chim. Acta*, in press.
(10) O. Sinanoglu, *Adv. Chem. Phys.*, **12**, 283 (1967).
(11) M. J. D. Powell, *Comput. J.*, **1**, 303 (1965).
(12) N. A. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys.*, **21**, 1087 (1953).
(13) K. Morokuma, *J. Chem. Phys.*, **55**, 1327 (1971); S. Iwata and K. Morokuma, *Chem. Phys. Lett.*, **19**, 94 (1974).
(14) A. Johansson and P. A. Kollman, *J. Am. Chem. Soc.*, **94**, 6196 (1972).
(15) J. E. Del Bene, *J. Am. Chem. Soc.*, **95**, 6517 (1973).
(16) W. J. Hehre, R. Ditchfield, and J. A. Pople, *J. Chem. Phys.*, **56**, 2257 (1972).

Ground States of Molecules. 44.[1] MINDO/3 Calculations of Absolute Heat Capacities and Entropies of Molecules without Internal Rotations

Michael J. S. Dewar* and George P. Ford[2]

Contribution from the Department of Chemistry, The University of Texas at Austin, Austin, Texas 78712. *Received January 21, 1977*

Abstract: MINDO/3 calculations of $S°$ and $C_p°$ are reported for 27 molecules in which internal rotations are absent. Agreement between these data and corresponding quantities calculated from spectral data is excellent, being in almost every case within one entropy unit at 298.15 K. Entropies of activation for six reactions are calculated and found to be in good agreement with experiment. These results suggest that such calculations may assist in studies of reaction mechanisms and equilibria involving species whole vibrational frequencies cannot easily be determined by experiment.

Introduction

Thermodynamic quantities are now routinely calculated from vibrational frequencies and moments of inertia when these are know from experiment.[3] Indeed the success of such calculations has led almost to the extinction of thermochemical measurements in this connection. Similar calculations could of course also be carried out using vibration frequencies determined by quantum mechanical methods. The results of such treatments are, however, in general both less reliable and less accurate than those given by experiment. Nevertheless there are important situations where such experimental data are lacking and where theoretical estimates of vibration frequencies could be very useful, faute de mieux. Three obvious examples are entropies of activation, equilibria involving transient species at high temperatures, and the thermodynamic properties of large molecules. The vibrational frequencies of transition states cannot in principle be measured experimentally, while technical difficulties often prevent such information being obtained for transient intermediates. In the case of large